SYNSENSO

Cell-free synthetic biology for combinatorial biosensor design

Project no. 101072980

Deliverable D1.1
Characterized RNA regulators

Version 1.2

WP1 – Building Cell-Free Computing Circuits

Lead Participant	Eindhoven University of Technology (TUE)
Contributors	Tom De Greef (TUE), Gökçe Özkul (TUE), Angelina Yurchenko (TUE)
Delivery date	29 February 2024
Revision date	29 January 2025
Dissemination level	Public
Type	Report

Funded by the European Union

Engineering and Physical Sciences Research Council

Funded by the European Union (GA 101072980) and supported by the UK Engineering and Physical Sciences Research Council. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Revision history

Date	Authors	Revision
29/01/2024	Tom F.A. de Greef, Gökçe Özkul, Angelina Yurchenko	Draft version
02/02/2024	Tom F.A. de Greef	First version
13/02/2024	Heinz Koeppl, Christiane Hübner	Internal review
29/02/2024	Heinz Koeppl	Final version
09/01/2025	Tom F.A. de Greef	Revised version
29/01/2025	Heinz Koeppl, Christiane Hübner	Final revised version

Contents

Abbreviations

Abbreviations	Details
BO	Bayesian optimization
GFP	Green fluorescent protein
OD	Optical density

Partner Short Names

Abbreviations	Details
TUDa	Technische Universität Darmstadt
TUE	Eindhoven University of Technology
IIT	Istituto Italiano di Technologia
TVU	Tor Vergata Universita degli Studi di Roma
DBS	Dynamic Biosensors GmbH
UBI	Ulisse Biomed S.p.A.
ICL	Imperial College London
ABV	Abvance Biotech S.L.
ACC	accelopment Schweiz AG
LEB	LenioBio GmbH
BRA	BRAIN Biotech AG
NUC	Nuclera Nucleics Ltd.
BME	Biomerieux SA
UNA	University of Naples Federico II

Executive Summary

This document reports on deliverable D1.1, “Characterized RNA regulators”. This document outlines the scientific progress in WP1:

- Production and optimization of Escherichia coli-based cell lysates for transcription/translation (TX-TL) systems.
- Development of cloning strategies for CRISPR-Cas9-based riboregulators.
- Preliminary integration of machine learning (ML) for system optimization.

This Deliverable is a plan on WP1, Objective 1 and 2.

Objectives of the Deliverable

This deliverable describes the production, characterization and data-driven optimization of cell lysate and the development of riboregulators for use in cell-free gene expression. This work was conducted by DC3 and DC4 and is part of WP1 of SYNSENSO.

The optimized production and detailed characterisation of the cell lysate allows to:

- Identify, design, and optimize regulatory elements for use in cell-free systems.
- Build Boolean-type signal integration circuits that are composable, have low leakage and fast response times.

Outcomes

We succeeded in preparing functional cell-lysate at the TU/e as evidenced by successful production of green fluorescent protein (GFP). In addition, we have started cloning plasmids that will allow cell-free expression of CRISPR-Cas-based riboregulators. Finally, we have developed a machine learning model that allows data-driven optimization of cell-free gene expression.

Next steps

In the next few months, we will establish cell-free gene expression of CRISPR-Cas9 based riboregulators. We will screen various conditions by employing high-throughput screening in combination with machine learning to find conditions that allow for maximal expression of the riboregulators.

1 Production of cell lysate

In this section, we describe the production and optimization of cell lysate. Cell lysate has been characterised using plasmids that produce GFP.

1.1 Cell-lysate production at the TU/e

At the TU/e we succeeded in producing cell-lysate using protocols outlined in literature.[1] In short, we prepared cell-lysate based on *Escherichia coli* Rosetta2 cells. The cells are grown initially on an LB-agar growth media, and then transferred to and scaled up in liquid LB growth media successively. The cells were collected as they reached a certain growth state verified by Optical Density (OD) measurements. Collected cell pellet was treated and prepared for lysate preparation as described in literature.[1] Cell-lysate production was achieved by sonication following the provided specifications by Kwon and Jewett.[2] The lysates were optimized to enhance the presence of critical cofactors and ensure sufficient translation activity. Protocols were iteratively refined based on OD measurements and GFP output. This foundation is critical for ensuring high fidelity in downstream applications, including RNA regulator assays

1.2 Characterisation of cell-lysate

To test the functionality of the prepared cell-lysate we used a plasmid (Figure 1a) that expresses GFP upon mixing with the cell-lysate. GFP expression is achieved through a gene cassette including the intrinsic σ70 promoter of *E. coli*. We measured the fluorescence in a plate reader for 16 hours. Our results show that the prepared cell-lysate functions well as evident by the production of GFP. In addition, the expression of GFP is compared to another cell-free protein expression system that is known to work (Figure 1b). The cell-lysate we prepared, i.e. light green in Figure 1b, performs similar to a verified cell-free expression system [3], that is given in dark green in Figure 1b. Non-template control, given in grey in Figure 1b, was also included to indicate the fluorescence is obtained solely due to the plasmid provided to the cell-free expression system. Fluorescence analysis using a GFP-expressing plasmid confirmed the functionality of the optimized lysates. Comparisons against benchmarked cell-free systems demonstrated comparable performance.

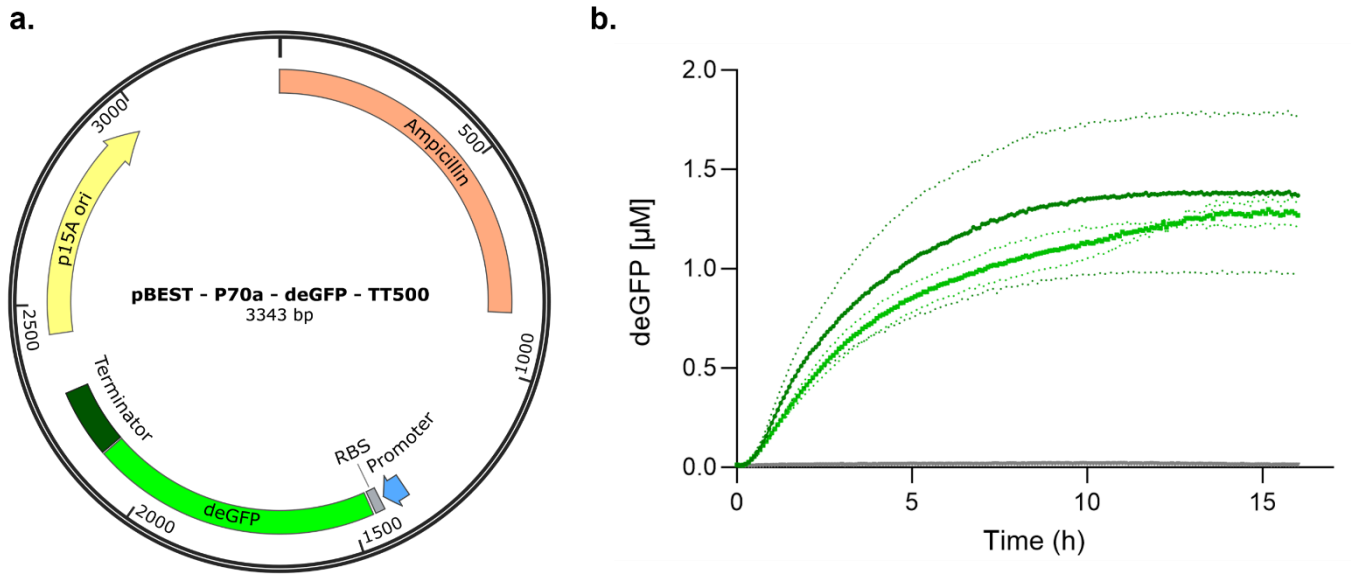

Figure 1: a, Plasmid provided to the cell-free reaction mixture, i.e. cell lysate mixed with feeding buffer, for GFP expression. The expression is under the control of σ70 promoter, and stopped by T500 Terminator. The plasmid is also used to prepare the linear DNA templates for other trials. b, performance of the cell-lysate was tested using the plasmid provided in a and shown in light green. The performance of the cell-lysate was compared to that of a well characterized cell-free expression system displayed by Bartelds et.al, shown in dark green. The data also includes a non-template control given in grey, showing the cell-free expression system works only in the presence of a DNA template.

2 Cloning of RNA regulators

2.1 Cloning Strategy

At the TU/e, we first designed constructs for the integration of Cas9 to our cell-free protein expression system. Our aim is to understand the activity and functionality of a well-known CRISPR-Cas effector protein within our system. The designed templates for Cas9 expression and sgRNA expression will be cloned using Gibson Assembly. This technique specializes on combining gene fragment(s) with a backbone vector in terms of complementary overhangs. The integration and functionality of Cas9 will be detected as a low expression of deGFP due to sgRNAs designed to target promoter and deGFP regions of the vector introduced to the cell-free expression system (Figure 2). For the activities described in D1.1, the focus has been on CRISPR-Cas9 due to its well-documented performance in cell-free systems. While Cas7-11 is an interesting alternative, its characterization is planned in later project stages. We confirm that exploring Cas9 remains an objective, forming a critical pathway to achieving deliverable goals.

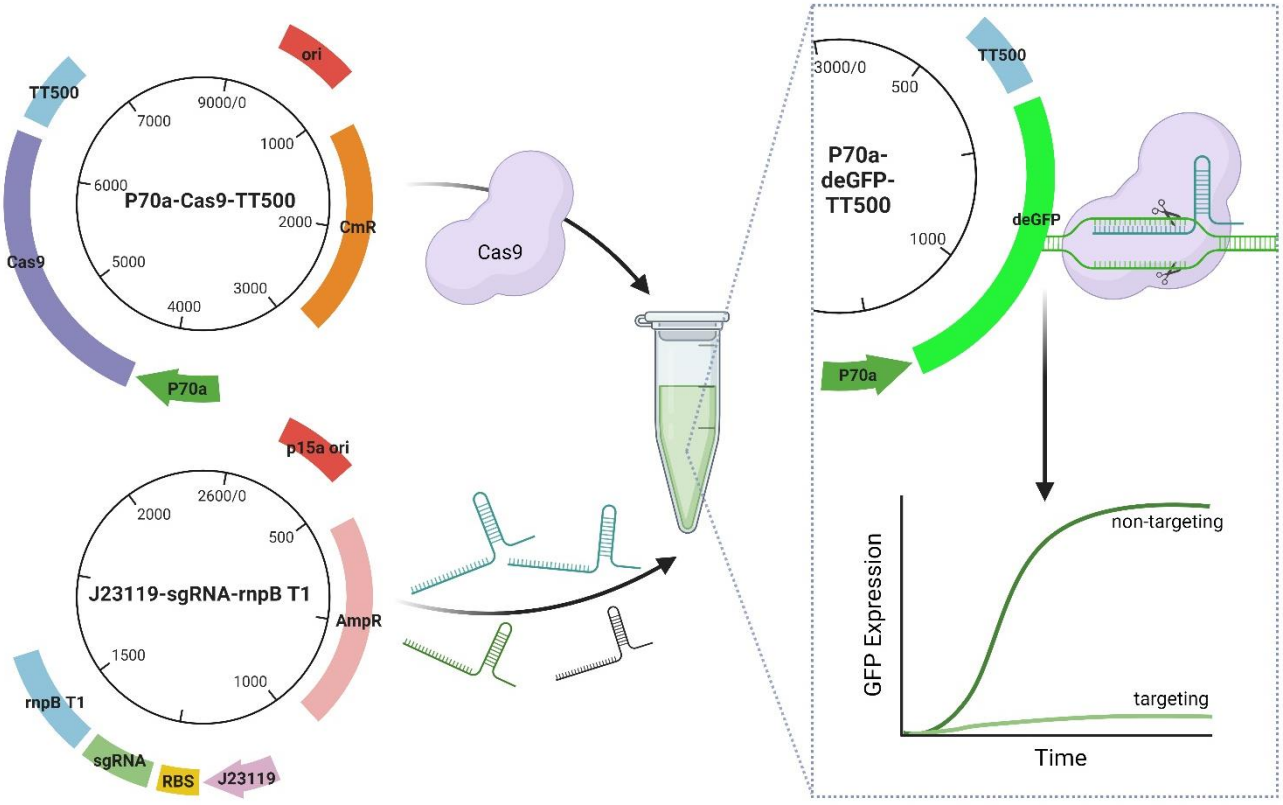

Figure 2: The integration of Cas9 into cell-free expression system. The gene cassettes are designed specifically for Cas9 (P70a – Cas9 – TT500) and sgRNA (J23119 – sgRNA – rnpB T1) expression in the cell-free system. sgRNAs are designed to interfere with deGFP expression that is encoded by P70a – deGFP – TT500 vector that is introduced to the cell-free expression system. Successful results will indicate a drop in the deGFP expression, which is measured and displayed in terms of fluorescence.

3 Machine learning model for optimization of cell-lysate

In this section, we describe a machine learning model for optimization of cell-lysate.

3.1 Machine learning model

For the optimization of the cell lysate system, we use the Bayesian optimization (BO) (Figure 3) that strategically selects the most informative data points for labeling, thereby reducing the necessary number of experiments to optimize the biological objective function. A machine learning model serves as a mathematical representation of a biological objective function. As a machine learning algorithm, we use Gaussian process regression which predicts yield values with corresponding uncertainties that are required for Bayesian Optimization.

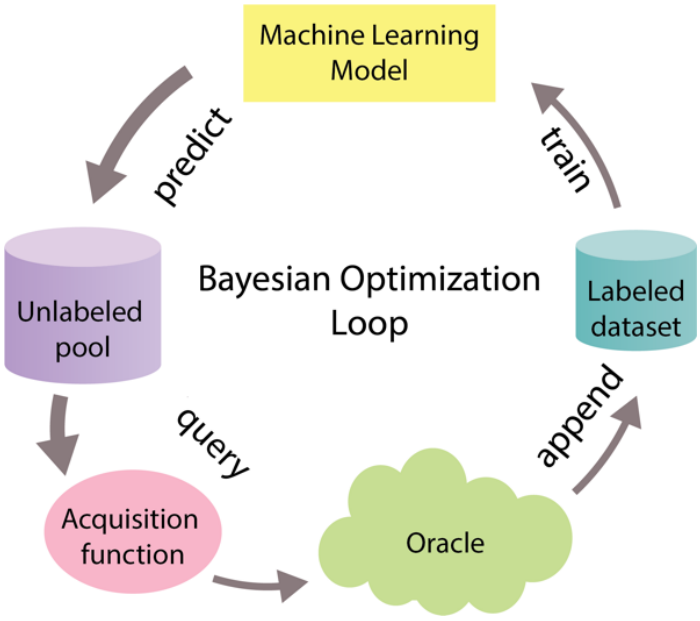

Figure 3: Bayesian Optimization algorithm. The Bayesian Optimization cycle initiates with the training of a machine learning model using an initial labeled dataset. Following this training, the model predicts labels and assesses the associated uncertainty for data points within an unlabeled pool. These predictions and uncertainty scores serve as inputs for the acquisition function, which strategically chooses the most informative data points to include in the sample. Subsequently, these highly informative data points are labeled with the assistance of an oracle (i.e. an experimentalist) and then incorporated into the initial labeled dataset. This iterative process continues until the desired outcome is achieved. Different widths of arrows represent the amount of data that goes through the cycle.

3.2 Minimum sample size required for optimization

To determine the minimum sample size required for optimizing cell-free GFP expression, we conducted Bayesian optimization with 5, 10, 25, 50, and 100 data points per round over a total of 10 rounds. Following the procedure outlined in Pandi et al. [4], we employed the golden regressor (i.e. a pre-trained model on the data from Borkowski et al. [5]) to label the most informative data suggested by a machine learning model. Our findings indicate that a minimum of 25 data points is necessary to effectively optimize the system and obtain a reliable predictive model (Figure 4).

4 Integration of ML with RNA Regulators

Machine learning is envisioned to play a pivotal role in characterizing RNA regulators by optimizing experimental designs for assessing Cas9 and sgRNA interactions. For this, the optimized model for GFP expression will be used as a starting point. Future iterations will incorporate active learning strategies to refine sgRNA designs for maximal regulatory efficiency. This deliverable demonstrates the initial steps toward developing characterized RNA regulators. The work performed—including lysate optimization, preliminary cloning, and ML integration—establishes a robust foundation for achieving the stated goals.

Additional focus on CRISPR-Cas9 characterization and its synergy with machine learning will ensure future milestones are met.

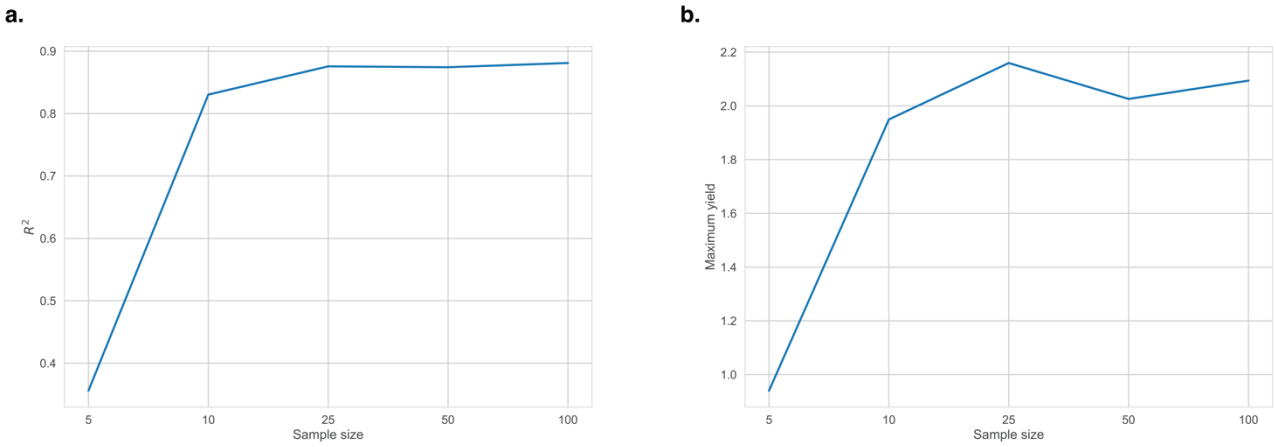

Figure 4: Minimum sample size required for optimizing cell-free GFP expression. a) Influence of the sample size on model accuracy (R^2). Model accuracy shows a significant increase from sample size 5 to 25, plateauing thereafter. b) Influence of the sample size on the maximum yield obtained after optimization. The highest yield is achieved with a sample size of 25 data points.

REFERENCES

1) Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., Noireaux, V. Protocols for Implementing an *Escherichia coli* Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp. 79*, e50762 (2013).
2) Kwon, YC., Jewett, M. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep* **5**, 8663 (2015).
3) Bartelds, M.W., García-Blay,O, Verhagen, P. G. A., Wubbolts, E. J., van Sluijs, B., Heus, H.A., de Greef, T.F.A., Huck, W. T. S., Hansen, M. M. K. Noise minimization in cell-free gene expression. *ACS Synth. Biol. 12,* 2217 (2023).
4) Pandi, A. et al. A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat. Commun. 13*, 3876 (2022).
5) Borkowski, O. et al. Large scale active-learning-guided exploration for in vitro protein production optimization. *Nat. Commun*. *11*, 1872 (2020).